© DEC 2022 | IRE Journals | Volume 6 Issue 6 | ISSN: 2456-8880

# An In-Depth Review on COVID-19 Infection Count Prediction Using Deep Learning

MAITHALI S. PATIL[1], PROF. BHAGYASHREE DHAKULKAR[2]
*[1, 2] Computer Dept., Dr. D. Y. Patil SOET, Lohegaon, Pune*

***Abstract- The COVID-19 pandemic has been one of the most defining factor of the recent decade due to its extreme impact on the world and the people within it. The coronavirus caused a lot of problems at the personal level as well as the economic level. There have been massive losses of life and the governments across the world have been hit very badly and are yet to recover from this pandemic. The main problem that has led to the escalation of the pandemic has been the fact that it was unexpected. The lack of highly accurate prediction approaches led to the medical infrastructure not being able to handle the load of patients. Therefore, there is a need for a methodology that can predict the COVID patient count to adequately prepare the healthcare professionals in advance. The proposed methodology utilizes deep learning in the form of Convolutional Neural Networks and Long Short Term Memory to achieve the prediction of COVID-19 infection count. This approach will be elaborated further in the upcoming editions of this research.***

***Indexed Terms- Convolutional Neural Networks-Long Short Term Memory, Decision Tree.***

## I. INTRODUCTION

Businesses, organizations, universities, and most crucially governments have begun to place a premium on data analysis and forecasting as a result of recent developments in digital communication and computers, as well as the rise of information analytics. Using complex mathematical methods has historically drawn the attention of academics, and understanding the features of an epidemic is no exception. As a consequence of the ongoing coronavirus pandemic, several mathematical methods are being investigated to better predict the epidemic's propensity, estimate the maximum, and compute its advancement.

The coronavirus epidemic, sometimes called coronavirus disease, is one of the world's most pressing problems. This novel and exceedingly contagious respiratory ailment has its origins in a coronavirus. Wuhan, China is credited as the place where the first mention of the formation of coronavirus was made; since then, the disease has swiftly spread and impacted societies all over the world. Though the high death toll from coronavirus is alarming, the problem persists. Coronavirus disease, caused by the SARS-CoV-2 virus, is a severe respiratory ailment.

Public health activities and demand forecasts are essential for recognizing and managing health risks. These efforts prioritize prevention above assessment, detection, and response to incidents, especially pandemics of infectious diseases. In order to create efficient preventive measures, it is necessary to have a solid estimate of the disease's current location, its rate of spread, its likely course, and the number of cases. Therefore, especially for infectious diseases, accurate quantitative predictive evaluation is an essential feature of public health efforts. European countries made little progress in stopping the spread of the coronavirus pandemic because their predictions were so off. Since coronaviruses cannot be halted at their origin, they have spread all across the globe.

The impact of coronavirus on the world's economy and population has been unprecedented. A vast number of people were infected, and numerous people lost their lives as the pandemic rapidly spread over the globe with no signs of slowing down. Combating coronavirus and future pandemics requires the development and implementation of reliable management systems. The development of such strategies requires precise tracking of the spatial and temporal development of sickness. Therefore, the powers of projected surveillance and the development of methods for epidemic containment might be considerably enhanced by a reliable and broadly comprehensible method of forecasting the virus's propagation.

Understanding the behavior of the pandemic is vital for beginning to effectively prepare health care professionals and other associated organizations. Simulation and forecasting of coronaviruses are of highest importance for the delivery of preventative services. To slow down the transmission of the coronavirus, sophisticated analytics are required. The government can better prepare for future issues similar to an epidemic if it knows how the virus would spread, in terms of both manpower and medical facilities.

To remedy the major shortcomings and issues raised by existing COVID- 19 patient contact tracing systems, Charuka Moremada et al. [1] introduce a contact tracing system that is equipped with BLE and GPS capabilities. As an added bonus, this system may be used both online and offline, offering administrators and end users a lot of options. Additionally, the writers could assess and generate infection probabilities for users with the help of an implemented algorithm. In addition, the researchers used a graph-based technique to simulate the algorithms and compared the results to real data collected in various places.

In order to anticipate future expansion of infected cases, Ankit Ramchandani et al. [2] offer a unique deep learning model to analyse heterogeneous variables at the county level. The suggested technique utilises the temporal and spatial (if available) structure of the data to extract embeddings from multivariate time series and multivariate spatial time series in a unique manner. In order to analyse data of a similar kind, further deep learning studies might use this method of embedding extraction. In addition, the suggested model, in contrast to the previous models, accepts a wide number of input characteristics and learns relationships between them. The model's usefulness was proved by its ability to foresee the rise in the number of new cases in counties throughout the United States during the 2009 COVID-19 epidemic. The model's prediction ability is adequate, and the findings of its feature analysis are easily interpretable, giving policymakers a tool with which to combat the spread.

The AutoML framework of auto-sklearn was modified by Jaco Tetteroo et al. [3] to provide predictions for the COVID-19 outbreak. We utilised data on deaths and moves made by people in 26 European nations to build ensembles of regression models with automated configuration. We compared the efficiency of a multi-output ensemble to that of a repeating single-output ensemble, and we coupled the two with idea drift adaption techniques for future study. Through root-mean-squared error comparisons, we analysed how well our ensembles performed in comparison to five independent benchmarks taken from the most up-to-date COVID-19 forecasting literature. Our study has shown the promise of combining open mobility data and AutoML systems for COVID-19 forecasting.

This literature survey article divides section 2 into an assessment of prior work in the form of a literature review, and section 3 concludes with recommendations for future research.

## II. RELATED WORKS

To forecast and assess future COVID-19 fatalities, Onder Tutsoy et al. [4] suggested the SpID-N model. We have first established the internal and linked structure of the SpID-N model, and the authors have provided a keen study of the widely-used SIR model. Second, the mathematic models of non-pharmacological policies have been created and described in terms of known bases and their unknown contributions to each section of the SpID-N model. Finally, the authors develop an LS-based optimization strategy for estimating the SpID-N model's missing parameters. Two of the modes are near to the instability zone, however the anticipated model parameters show that the COVID-19 deaths in Turkey are inside the stable range.

IT-GCN was created by Zehua Yu et al. [5] who analysed the distribution of COVID-19 using it. In lieu of a geographical distance between nodes, an interaction temporal graph uses the Euclidean distance of parameter vectors to measure proximity. COVID-19 study shows that our suggested method can accurately forecast new instances of infection in the United States. Additionally, these forecasts may help with bettering containment practises. Results from the experimental study can be accurately quantified for the whole United States and the state of California in particular.

It has been reported by R. G. Babukarthik et al. [6] that the exponential rise in the number of COVID-19 patients has led to a global pandemic. The therapy of COVID-19 should be quicker and cheaper if the condition can be diagnosed early. Specifically, CXR image samples are utilised for COVID-19 prediction using a deep learning approach. While some individuals have contracted pneumonia in the actual world, the vast majority have not been afflicted. Therefore, there is a discrepancy between the sick individual and a healthy control when predicting pneumonia. Through the use of CXR image samples, the GDCNNN approach is suggested in this study for differentiating between COVID-19 and a healthy individual.

A validated RF-SMA-SVM model was created to identify the severity of COVID-19 based on criteria such as basic patient information and 26 blood routine indicators, as provided by Peiliang Wu et al. [7]. The two most novel aspects of this article are the first use of the SMA to train an optimum SVM classifier and the first proposal that the severity of COVID-19 may be discriminated using a combination of age, N/L, and P/L indicators. The experimental findings show that the suggested technique outperforms the SVM method based on other optimization algorithms in terms of prediction accuracy and stability when applied to the COVID-19 severity prediction issue, all while filtering out the important parameters rich in discriminating power.

High performance computing is very useful for obtaining model solutions and fine-tuning system parameters, as discussed by Vasyl Martsenyuk et al. [8]. This data was saved on the COVID data hub for the purpose of SciML. Using Julia parallel computing in R with the help of the diffeqr package, we can perform the fundamental tasks of integrating delayed differential equations 14 times quicker. The research introduces a model-based, delayed-dynamical-system-based Big Data best practise for epidemiological analytics. It was accomplished as part of the project iBigWorld, whose goal is to provide its participants with the knowledge and tools necessary to create novel solutions using Big Data in practical settings. Gaining experience with COVID Big Data is a great way to acquire skills in the latest Big Data and machine learning-based technologies.

The SEIHR compartmental model was introduced by Ruiwu Niu et al. [9] as a deterministic framework. The authors found a mathematical equation for the fundamental reproduction number using the latest iteration of the matrix approach. Researchers were able to generate parameter areas that allow for control of the epidemic breakout by evaluating the link between the reproduction number and the model parameters. It also discovered that enabling individuals to easily enter or exit the system might actually slow the pace of epidemic, at the cost of exporting illnesses to other places, when the local pandemic is more severe than the global one. The authors then conducted many simulations to demonstrate that the isolation of symptomatic infected individuals makes the epidemic process more sensitive to the transmission rate of the exposed than that of the infected.

In order to analyse the behaviour of the COVID-19 pandemic, Kartik Narayan et al. [10] use SIR, a feedback technique based on control theory and regression models. It was concluded that the SIR model is superior than the regression model during the outbreak period but superior to the regression model when assessing long-term patterns in the transmission of illnesses. Mean squared error between anticipated and observed values provides a quantitative measure of performance. The data presented in the research demonstrate a significant discrepancy between the expected and observed epidemic's inflection points early on. The graphic result and the mean squared error indicate that the SIR model is not suitable for making an early forecast of the outbreak in this situation.

In light of the present epidemic, Ertugrul Karaçuha et al. [11] provide concrete recommendations for estimating future caseloads in order to properly allocate healthcare resources for patients. This simulation is applicable not just to the current situation with COVID-19 but also to potential future regional or global outbreaks. In this work, researchers model and anticipate the amount of COVID-19-related confirmed cases, fatalities, and recoveries across many nations. To accomplish their objectives, the authors used an approach that makes use of deep evaluation and fractional calculus. Extensive experimental evaluations have been implemented, and the results reveal that the method is effective at producing accurate predictions and models with little room for mistake.

The SIR model used to estimate COVID-19 deaths was evaluated by Onder Tutsoy et al. [12]. The innovative, all-encompassing SpID model has then been presented, dissected, and defended. Subsequently, the COVID-19 deaths in Turkey were modelled, and the unknown parameters were identified using an LS based parametric optimization strategy. The proposed model successfully predicts the number of casualties in Turkey. Furthermore, the model estimates that the lowest possible number of infected and dead persons would be reached in 300 days, while the lowest possible number of suspicious casualties will be reached around 1000 days. Although the model is developed and evaluated using data from Turkey's COVID-19 victims, its main parameters may be modified to account for victims in other nations.

The authors Furqan Rustam et al. [13] warn that the vulnerability of the COVID-19 pandemic might spark a major international disaster. There is concern among some scientists and government bodies throughout the globe that the epidemic might have widespread effects. In this research, we present a machine learning-based approach for estimating the potential for a worldwide epidemic of COVID-19. Using machine learning techniques, the system examines a dataset that contains the real historical data broken down by day and predicts what will happen on future days based on that analysis. The study's findings demonstrate that, given the present state of the art and the type and amount of the dataset, ES provides the greatest performance in the predicting domain.

In a paper published in 2014, Lin Wang et al. [14] introduced the Lasso logistic regression model, advanced features concerning the pattern changes were extracted from the time-lapse profile of vital signs, and the concept of SpO2 "Bad Days" was introduced on the basis of the daily range of SpO2 readings. The results of this study show that patterns in patients' oxygen saturation (SpO2) not only helped to distinguish between those who had died and those who had been released, but also hinted at the possibility of using the model in clinical decision making about ICU admissions. To forecast the fate of COVID-19 patients, many machine learning algorithms have been developed. While these models have been useful, they may not be able to reliably forecast early COVID-19 mortality since their primary characteristics are laboratory values. Instead, this approach is built on the first 24 hours of essential data, allowing for prompt and effective therapeutic actions to be determined and implemented.

In order to estimate the number of confirmed cases, deaths, and recoveries from COVID-19 throughout India's several states, Vishan Kumar Gupta et al. [15] investigate five machine learning models with these three key characteristics. Confirmed cases, deaths, and recoveries are the qualitative indicators. In this scenario, machine learning approaches do not include any supplementary data from other models or template structures. The accuracy of each model is measured against the others. Based on the results of our extensive trials, we decided to use the random forest approach as our final prediction model for predicting the outcomes of our different scenarios. The reliability of the random forest model, which gave approximately linear performance to the prediction of all these situations, is evaluated using K-fold cross validation.

## III. PROPOSED METHODOLOGY

The presented approach has been realized through thorough analysis of the present techniques and the flowchart for the same is given below. At first, the start state is initiated and the dataset is provided as an input is preprocessed and labeled. The preprocessed and labeled data is subjected to correlation analysis through Pearson correlation and the data is clustered using K Nearest Neighbor clustering approach. The clusters are then utilized for LSTM models such as LSTM, BD-LSTM and CNN-LSTM. The resultant values are then classified using decision making for COVID-19 patient count prediction and the stop state is reached.

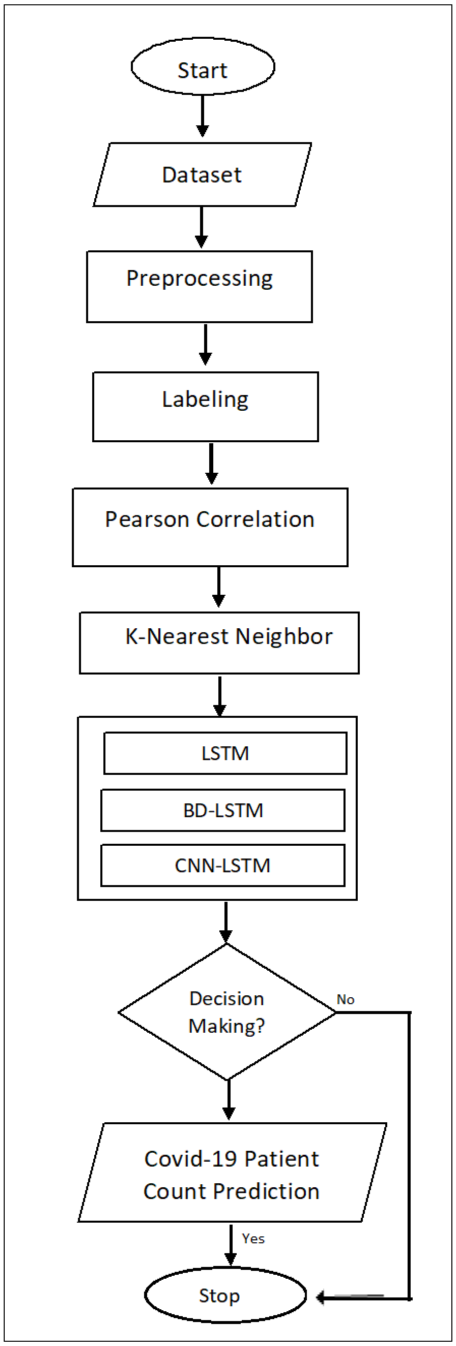

Figure 1: Flowchart for the proposed methodology

## CONCLUSION AND FUTURESCOPE

When considering the last decade and its significance, the COVID-19 epidemic stands out as one of the most influential events. The coronavirus was the root of many troubles, both private and monetary. Death tolls are in the millions, and governments throughout the globe have taken a severe financial blow from this epidemic. The surprise nature of the epidemic has been the primary contributing factor to its rapid spread. Because of a deficiency of extremely accurate prediction methods, the nation's healthcare system was overwhelmed and unable to treat all of its patients. In order to appropriately educate healthcare personnel in advance, a technique is required to anticipate the COVID patient count. To forecast the number of COVID-19 infections, the suggested technique employs deep learning in the form of Convolutional Neural Networks and Long Short Term Memory. Future iterations of this study will provide further detail on this strategy.

## REFERENCES

[1] C. Moremada, C. Sandeepa, N. Dissanayaka, T. Gamage and M. Liyanage, "Energy efficient contact tracing and social interaction based patient prediction system for COVID-19 pandemic," in Journal of Communications and Networks, vol. 23, no. 5, pp. 390-407, Oct. 2021, doi: 10.23919/JCN.2021.000037.

[2] A. Ramchandani, C. Fan and A. Mostafavi, "DeepCOVIDNet: An Interpretable Deep Learning Model for Predictive Surveillance of COVID-19 Using Heterogeneous Features and Their Interactions," in IEEE Access, vol. 8, pp. 159915-159930, 2020, doi: 10.1109/ACCESS.2020.3019989.

[3] J. Tetteroo, M. Baratchi and H. H. Hoos, "Automated Machine Learning for COVID-19 Forecasting," in IEEE Access, vol. 10, pp. 94718-94737, 2022, doi: 10.1109/ACCESS.2022.3202220.

[4] O. Tutsoy, A. Polat, Ş. Çolak and K. Balikci, "Development of a Multi-Dimensional Parametric Model With Non-Pharmacological Policies for Predicting the COVID-19 Pandemic

Casualties," in IEEE Access, vol. 8, pp. 225272-225283, 2020, doi: 10.1109/ACCESS.2020.3044929.

[5] Z. Yu, X. Zheng, Z. Yang, B. Lu, X. Li and M. Fu, "Interaction-Temporal GCN: A Hybrid Deep Framework For Covid-19 Pandemic Analysis," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 2, pp. 97-103, 2021, doi: 10.1109/OJEMB.2021.3063890.

[6] R. G. Babukarthik, V. A. K. Adiga, G. Sambasivam, D. Chandramohan and J. Amudhavel, "Prediction of COVID-19 Using Genetic Deep Learning Convolutional Neural Network (GDCNN)," in IEEE Access, vol. 8, pp. 177647-177666, 2020, doi: 10.1109/ACCESS.2020.3025164.

[7] P. Wu et al., "An Effective Machine Learning Approach for Identifying Non-Severe and Severe Coronavirus Disease 2019 Patients in a Rural Chinese Population: The Wenzhou Retrospective Study," in IEEE Access, vol. 9, pp. 45486-45503, 2021, doi: 10.1109/ACCESS.2021.3067311.

[8] V. Martsenyuk, M. Bernas and A. Klos-Witkowska, "Two-Strain COVID-19 Model Using Delayed Dynamic System and Big Data," in IEEE Access, vol. 9, pp. 113866-113878, 2021, doi: 10.1109/ACCESS.2021.3104519.

[9] R. Niu, E. W. M. Wong, Y. -C. Chan, M. A. Van Wyk and G. Chen, "Modeling the COVID-19 Pandemic Using an SEIHR Model With Human Migration," in IEEE Access, vol. 8, pp. 195503-195514, 2020, doi: 10.1109/ACCESS.2020.3032584.

[10] K. Narayan, H. Rathore and F. Znidi, "Using Epidemic Modeling, Machine Learning and Control Feedback Strategy for Policy Management of COVID-19," in IEEE Access, vol. 10, pp. 98244-98258, 2022, doi: 10.1109/ACCESS.2022.3206790.

[11] E. Karaçuha et al., "Modeling and Prediction of the Covid-19 Cases With Deep Assessment Methodology and Fractional Calculus," in IEEE Access, vol. 8, pp. 164012-164034, 2020, doi: 10.1109/ACCESS.2020.3021952.

[12] O. Tutsoy, Ş. Çolak, A. Polat and K. Balikci, "A Novel Parametric Model for the Prediction and Analysis of the COVID-19 Casualties," in IEEE Access, vol. 8, pp. 193898-193906, 2020, doi: 10.1109/ACCESS.2020.3033146.

[13] F. Rustam et al., "COVID-19 Future Forecasting Using Supervised Machine Learning Models," in IEEE Access, vol. 8, pp. 101489-101499, 2020, doi: 10.1109/ACCESS.2020.2997311.

[14] L. Wang et al., "A Time-Series Feature-Based Recursive Classification Model to Optimize Treatment Strategies for Improving Outcomes and Resource Allocations of COVID-19 Patients," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 7, pp. 3323-3329, July 2022, doi: 10.1109/JBHI.2021.3139773.

[15] V. K. Gupta, A. Gupta, D. Kumar and A. Sardana, "Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model," in Big Data Mining and Analytics, vol. 4, no. 2, pp. 116-123, June 2021, doi: 10.26599/BDMA.2020.9020016.